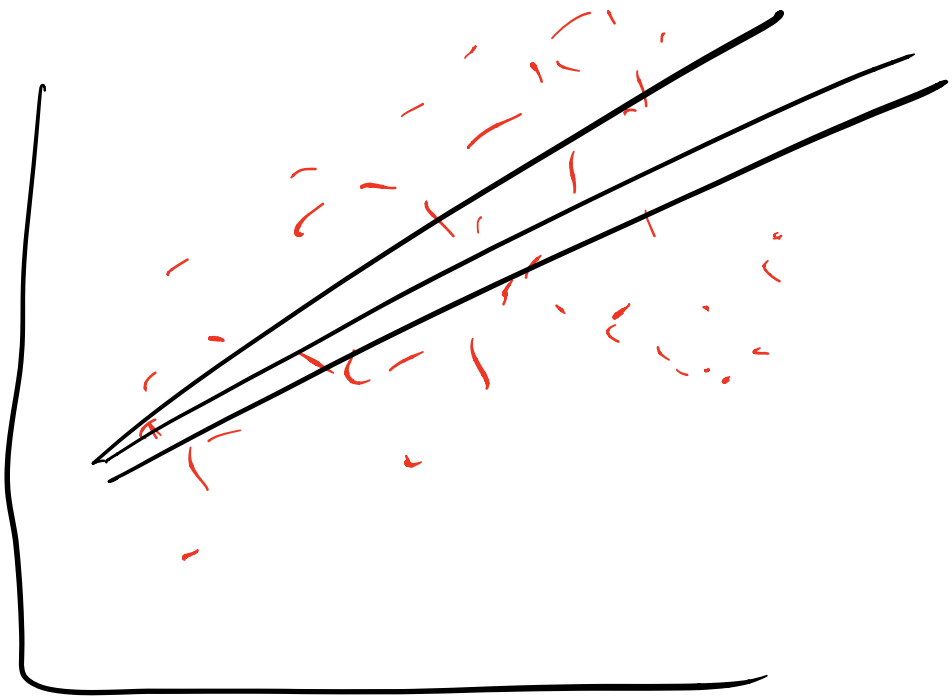
Why do we regression this way?



The regression line we use is the line with smallest rms error.

How to interpret slope?

$$y = mx + b$$

" an increase in x by 1, means we would predict an increase in y by m"

Correlation does not imply causation.

χ^2-test

"chi-squared"

a) Counting the number of times each category occurs (≥ 2)

b) Test for independence between

2 (qualitative) variables

a) 1) $\chi^2 = \sum_{\#\text{categories}} \frac{(obs - exp)^2}{exp}$

2) $df = \#cat - 1$

p-value: $1 - pchisq(stat, df)$

b) 1) $\chi^2 = \sum_{\#\text{categories in m x n table}} \frac{(obs - exp)^2}{exp}$

2) $df = (m-1)(n-1)$